



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **Recursive autoencoders based unsupervised feature learning for hyperspectral image classification**

Zhang, X., Liang, Y., Li, C., Jiao, L., & Zhou, H. (2017). Recursive autoencoders based unsupervised feature learning for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*.  
<https://doi.org/10.1109/LGRS.2017.2737823>

**Published in:**  
IEEE Geoscience and Remote Sensing Letters

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**  
© 2017 IEEE.  
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

**General rights**  
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

# Recursive Autoencoders Based Unsupervised Feature Learning for Hyperspectral Image Classification

Xiangrong Zhang, *Senior Member, IEEE*, Yanjie Liang, Chen Li, Ning Huyan, Licheng Jiao, *Senior Member, IEEE*, Huiyu Zhou

**Abstract**—For hyperspectral image (HSI) classification, it is very important to learn effective features for the discrimination purpose. Meanwhile, the ability to combine spectral and spatial information together in a deep level is also important for feature learning. In this letter, we propose an unsupervised feature learning method for HSI classification, which is based on recursive autoencoders (RAE) network. RAE utilizes the spatial and spectral information and produces high-level features from the original data. It learns features from the neighborhood of the investigated pixel to represent the whole local homogeneous area of the image. In addition, to obtain more accurate representation of the investigated pixel, a weighting scheme is adopted based on the neighboring pixels, where the weights are determined by the spectral similarity between the neighboring pixels and the investigated pixel. The effectiveness of our method is evaluated by the experiments on two hyperspectral datasets and the results show that our proposed method has better performance.

**Index Terms**—unsupervised feature learning, recursive autoencoders, deep learning, hyperspectral image classification.

## I. INTRODUCTION

With the developments of remote sensing technologies, hyperspectral images (HSI) captured by hyperspectral imaging sensors have been successfully used to detect and classify objects. HSI consists of a number of bands even up to several hundreds over the spectrum, which includes a wealth of information, thus has been widely applied for many practical applications such as the management of the environment, precision agriculture, social security and military defense as well as mineralogy [1-3]. One of the most important applications of HSI is image classification. There are many available methods for HSI classification, which can be categorized into two types, one is pixel-based and the other is

spatial-spectral based. Due to the spatial structure information considered in the applications, promising performance has been achieved [4][5] by spatial-spectral methods.

In the past decades, many effective feature learning methods for HSI have also been proposed, such as principal component analysis [6], kernel principal component analysis [7], particle swarm optimization-based feature selection [8], and semi-supervised feature learning [9]. Recently, spatial-spectral feature learning methods have found their applications in HSI analysis, such as 3D Gabor filter [10], 3D gray-level co-occurrence [11] and composite kernel SVM [12]. In [13], a manifold learning method was used to improve the band selection quality. Sparse coding also can support feature learning [14][15] which aims to find few proper bases to represent the original information. In [16], sparse representation and graphical model are combined for HSI classification. To make full use of spatial information, in [17], a novel spectral-spatial classification method using multitask joint sparse representation and a step-wise Markov random field framework was proposed to further improve the classification accuracy.

For feature learning, many unsupervised feature learning methods based on deep learning have been proposed [18-20]. Recently, some deep learning methods have also been used to process hyperspectral images. In [21], Stacked Auto Encoders (SAE) was used for HSI classification, in which spatial information is used in the SAE model. The restricted Boltzmann machine (RBM) and Deep Belief Network (DBN) on hyperspectral images were reported in [22] and they designed a novel deep architecture to combine spectral and spatial information. Convolutional Neural Network (CNN) has also been used for hyperspectral feature extraction [23] for better results.

Semi-supervised recursive autoencoders (RAE) [24] based on recursive neural networks [25] has shown excellent performance for natural language processing. Inspired by semi-supervised RAE, we here propose an unsupervised RAE network with spectral-spatial learning for HSI classification. Firstly, an unsupervised RAE network is built by the training data, and then the network weights are trained using back propagation (BP) and the limited memory BFGS method (L-BFGS) [26]. Afterwards, the similarity weights between the investigated pixel and its neighboring pixels are integrated into the network, which determines the merging sequence and importance. Finally, the output of the RAE network represents higher-level features of the investigated pixel in the neighborhood, whose dimension is the same as the number of the spec-

This work was supported in part by the National Natural Science Foundation of China (nos. 61272282, 61377011 and 61373111), and the Program for New Scientific and Technological Star of Shaanxi Province (No. 2014KJXX-45). H. Zhou is supported by UK EPSRC under Grants EP/N508664/1, EP/R007187/1 and EP/N011074/1, and Royal Society-Newton Advanced Fellowship under Grant NA160342.

X. Zhang, Y. Liang, N. Huyan, and L.C. Jiao are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, P. R. China. (Email: xrzhang@ieee.org).

C. Li is with the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710048, P. R. China.

H. Zhou is with the School of Electronics, Electrical Engineering and Computer Science, Queen's University of Belfast, Belfast, BT7 1NN, UK.

tral bands. It should be noted that the process of feature extraction is unsupervised, thus this method overcomes the practical challenge when there are limited labeled samples available to use. Finally, the learned features are classified using support vector machine (SVM) which has achieved good performance on hyperspectral classification [27].

The contributions of the letter are two-fold: 1) a new unsupervised feature learning method based on recursive autoencoders is proposed for HSI classification; 2) a similarity weight between the investigated pixel and its neighboring pixels is considered in the RAE network to learn deep features.

## II. AUTOENCODERS

Autoencoders (AE) network [21] is a symmetrical neural network to learn features from the original data in an unsupervised manner. It is based on a single-hidden-layer feedforward neural network (SLFNN). Given the spectral vector of a pixel in HSI,  $\mathbf{x} \in \mathcal{R}^d$ , the hidden neurons in a SLFNN are denoted as  $\mathbf{h} \in \mathcal{R}^m$  ( $m$  is the number of hidden neutrals), which is achieved by:

$$\mathbf{h} = f(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) \quad (1)$$

where  $\mathbf{W}_1$  is an  $m \times d$  weight matrix,  $\mathbf{b}_1$  is a bias vector with  $m$ -dimension, and  $f$  is an activity function, which is a tanh function in this letter.  $f$  can also be a sigmoid function.

The AE network aims to learn an approximation to the identity function so that the output  $\hat{\mathbf{x}}$  of the reconstruction layer is similar to  $\mathbf{x}$ . The output neurons are defined as

$$\hat{\mathbf{x}} = \mathbf{W}_2 \mathbf{h} + \mathbf{b}_2 \quad (2)$$

where  $\mathbf{W}_2$  is a  $d \times m$  matrix and  $\mathbf{b}_2$  is the bias. The reconstruction error is

$$E = \frac{1}{2} \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \quad (3)$$

Minimizing the reconstruction error, the weight matrices  $\mathbf{W}_1$  and  $\mathbf{W}_2$ , and the biases  $\mathbf{b}_1$  and  $\mathbf{b}_2$  can be obtained via the standard BP algorithm. Then the outputs of the hidden neurons are the learned features of the original data.

The network is capable of carrying out dimensionality reduction through limiting the number of the hidden neurons, and it is proved that the hidden neurons are an excellent representation of the original data with less loss of information.

## III. UNSUPERVISED RECURSIVE AUTOENCODERS FOR FEATURE LEARNING

RAE network is a combination of the AE network and a recursive structure, which is to learn high-level features of the input as a representation [24]. The unsupervised RAE network based feature learning can overcome the difficulty of collecting labeled samples because of the unsupervised manner. In addition, the spatial-spectral information is also integrated in the unsupervised RAE network.

### A. Unsupervised RAE

RAE network adopts a recursive structure which is common, simple and accessible in practice. For example, in [25], a

building region can be recursively split into smaller regions depicting parts such as roofs and windows. Then we can merge smaller parts into a larger region according to the rule that the parts of the same class must firstly be merged and the ones of different classes should be merged later on. Finally, by the rule, a whole building merged by parts will be formed, standing for the original one, i.e. the entire building region. The recursive representation represents rationally the entire region and the dimension of the representation will not increase.

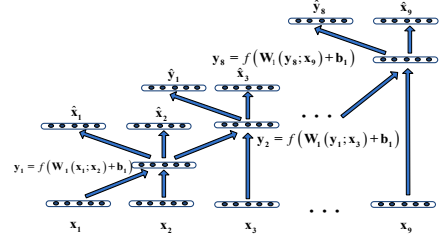


Fig. 1. The structure of unsupervised recursive autoencoders

The approach presented in [24] is used for image classification, and a supervised recursive structure is applied to learn features for each image. While our work is to generate the classification map of the land covers in HSI, we recursively learn features for each pixel based on an unsupervised recursive structure. To effectively use the recursive structure and take the spatial information into consideration as well, we propose a feature learning method using spatial neighborhood information in the unsupervised RAE network.

Unsupervised RAE is a tree structure. Taking a neighborhood region  $3 \times 3$  for example, the tree structure of unsupervised RAE is shown in Fig. 1. The structure is determined by the reconstruction error. The pixels in the neighborhood region are denoted as  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_9)$ ,  $\mathbf{x}_i \in \mathcal{R}^d$ . The form of branching triplets of the parent with children in the binary tree is represented as  $(\mathbf{p} \rightarrow \mathbf{c}_1, \mathbf{c}_2)$ , where  $\mathbf{p}$  denotes the parent and  $\mathbf{c}_1, \mathbf{c}_2$  are the children. For reconstructing the tree shown in Fig. 1, the representation of the tree form is:  $((\mathbf{y}_1 \rightarrow \mathbf{x}_1, \mathbf{x}_2), (\mathbf{y}_2 \rightarrow \mathbf{x}_3, \mathbf{x}_4), \dots, (\mathbf{y}_8 \rightarrow \mathbf{x}_9, \mathbf{x}_7))$ . To apply the same AE network to each pair of children recursively, the representations of hidden  $\mathbf{y}_i$  can keep the same dimensionality as  $\mathbf{x}_i$ .

The parent node  $\mathbf{p}$  (e.g.  $\mathbf{y}_1$ ) is computed from its children  $(\mathbf{c}_1; \mathbf{c}_2) = (\mathbf{x}_1; \mathbf{x}_2)$ :

$$\mathbf{p} = f(\mathbf{W}_1[\mathbf{c}_1; \mathbf{c}_2] + \mathbf{b}_1) \quad (4)$$

where the parameter matrix  $\mathbf{W}_1 \in \mathcal{R}^{d \times 2d}$  is multiplied by the concatenated vector of the two children  $[\mathbf{c}_1; \mathbf{c}_2]$ . After adding a bias term  $\mathbf{b}_1$ , we apply an element-wise activation function such as tanh or sigmoid to the resulting vector. In our work, the tanh activation function is used as same as the one used in [24]. Then the AE network is adopted to reconstruct the children in the reconstruction layer, which can assess how good the learned  $d$ -dimensional vector represents its children. The reconstructed children can be obtained by

$$[\hat{\mathbf{c}}_1; \hat{\mathbf{c}}_2] = \mathbf{W}_2 \mathbf{p} + \mathbf{b}_2 \quad (5)$$

In the AE network, the objective function for training is to minimize the reconstruction error of the input pairs. For each pair, the Euclidean distance between the original input and its reconstruction is computed to measure the reconstruction error:

$$E([\mathbf{c}_1; \mathbf{c}_2]) = \frac{1}{2} \|\mathbf{c}_1; \mathbf{c}_2] - [\hat{\mathbf{c}}_1; \hat{\mathbf{c}}_2]\|_2^2 \quad (6)$$

The above process depicts how good a  $d$ -dimensional parent vector representation  $\mathbf{P}$  is learned from two  $d$ -dimensional children  $(\mathbf{c}_1, \mathbf{c}_2)$ . The learned parent vector  $\mathbf{P}$  (e.g.  $\mathbf{y}_1$ ) is then treated as a child and combined with its nearest child (e.g.  $\mathbf{x}_3$ ) to learn new features (i.e.  $\mathbf{y}_2$ ) by AE network. By recursively learning features with AE network, a top parent node of the tree occurs, which is the representation of the investigated pixel in the spatial neighborhood.

However, here is a question: which pair of pixels should be first integrated? This can be solved through a rule that the neighboring pixels with the minimum reconstruction error have a higher priority for integration. As shown in Fig. 1, given the pixels in the neighborhood  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_9)$  as the input of the RAE network, the reconstruction error of the first pair of the neighboring pixels  $(\mathbf{x}_1; \mathbf{x}_2)$  is the lowest, so  $(\mathbf{x}_1; \mathbf{x}_2)$  are defined as the potential children of a patch  $(\mathbf{c}_1; \mathbf{c}_2) = (\mathbf{x}_1; \mathbf{x}_2)$  to the input of AE. The obtained parent  $\mathbf{y}_1$  will represent the pair of pixels  $(\mathbf{x}_1; \mathbf{x}_2)$ , and the input becomes  $(\mathbf{y}_1, \mathbf{x}_3, \dots, \mathbf{x}_9)$ . Then for each pair of children, we save the potential parent node  $\mathbf{P}$  and the obtained reconstruction error. Again, we find the pair of children owing the lowest reconstruction error and integrate them to get a new parent node. The process repeats until the last pair of children are merged into the top node, and the tree structure is built. So far, the unsupervised RAE network is constructed.

#### B. Learning Parameters

The model parameters  $\boldsymbol{\theta} = (\mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2)$  of the tree structure are learned by minimizing the reconstruction error shown in Eq. (6). The gradient on the reconstruction error becomes:

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = \frac{1}{N} \sum \frac{\partial E(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta} \quad (7)$$

To compute the gradient, all the candidate trees are first greedily constructed, and then the derivatives for these trees are computed via BP through structuring. A simple modification to the general BP used in RAE [23] is that the reconstruction errors are split at each node and then propagated to the children. Finally, L-BFGS is used over the training data to minimize the objective on the tree structure. Hence the model parameters  $\boldsymbol{\theta}$  is obtained and the unsupervised RAE network is built with the training samples.

#### C. Weighted Network

In our work, the goal of the unsupervised RAE network is to represent the investigated pixel using the learned features, but in a neighborhood, especially around the border of the regions, not all the neighboring pixels belong to the same class as the

central pixel. So, a weight determined by the similarity between the investigated pixel and the neighboring pixels is added to the network to effectively represent the investigated pixel. The larger the similarity weight is, the more important role the corresponding pixel plays in the unsupervised RAE network to learn features. The similarity weight is obtained by

$$S_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (8)$$

where  $\mathbf{x}_j$  is the investigated pixel and  $\mathbf{x}_i$  is a neighboring pixel.

Given the built tree structure and the weights, the leaf nodes  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_9)$  can be denoted as  $(S_{15}\mathbf{x}_1, S_{25}\mathbf{x}_2, \dots, \mathbf{x}_5, \dots, S_{95}\mathbf{x}_9)$  (assuming the investigated pixel is  $\mathbf{x}_5$ ). They are substituted into the trained tree by Eq. (4) to recursively obtain the top node. Then the average of all the nodes in the unsupervised RAE tree structure is used as the final features for the central pixel. Finally, SVM is selected as the classifier to assess the effectiveness of the learned features for HSI classification.

### IV. EXPERIMENT RESULTS

In this section, we evaluate our proposed weighted unsupervised RAE (unRAE) on two real HSI datasets and make comparison against four competitive classification algorithms: 1) Spectral-feature-based SVM classifier (SVM); 2) Spectral-Spatial feature based SAE [1] Spectral-spatial SAE); 3) Composite kernel SVM (SVM-CK) [12] and 4) joint sparsity model (SOMP) [14]. In Spectral-Spatial SAE, we use the window size of  $7 \times 7$  to collect the spatial information and reduce the dimension of features to 4 by PCA as suggested in the original paper. Only one hidden layer is used and the hidden units are set to half of the dimension of the original spectral features. For SVM-CK, the spatial information is provided by the extended multi-attribute profile (EMAP) [12]. The parameters of SVM-CK are obtained by fivefold cross-validation. For SOMP, the parameter settings are the same as that shown in [14], which applied spatial-spectral information to sparse representation for classification. The radial basis function (RBF) kernel is used in the SVM. For the proposed method, the spatial neighborhood of a square window is adopted for both the Indian Pines and the University of Pavia datasets.

#### A. AVIRIS Data Set: Indian Pines

The Indian Pines dataset was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Indiana Pine site area in Northwest Indiana, USA in June 1992. The AVIRIS sensor acquires 220 bands over the spectral range from 0.2 to 2.4  $\mu\text{m}$ . We remove 20 low signal-to-noise bands (104-108, 150-163 and 220) due to the water absorption, and the remaining 200 bands are used for the experimental analysis. The dataset consists of  $145 \times 145$  pixels with a spatial resolution of 20m/pixel, and has 16 classes of land covers available based on the ground truth information.

For all the algorithms, we randomly choose 10% of the labeled samples per class as the training set, and the remaining data is used as test samples. Table I shows the classification accuracy using different methods. Fig. 2 demonstrates the ground truth and the classification maps by different algorithms.

Table I shows that the proposed unRAE algorithm outperforms the other approaches in terms of overall accuracy and kappa coefficient. The pixel-wise SVM does not lead to better performance because no spatial structure feature is used. Although the spectral-spatial SAE [1] utilizes the spatial information the result is not improved because 10% of the labeled samples used in our paper are not sufficient for training the model. The SVM-CK, SOMP and our proposed method are compared here. For SOMP, it assumes that the HSI pixels in a small spatial neighborhood are approximated by a sparse linear combination of several atoms [14]. Compared with SOMP, unRAE-SVM obtains a better result in terms of overall accuracy and kappa coefficient due to the usage of a similarity weight. SVM-CK utilizes EMAP to obtain spatial information which improves the capability in extracting the spatial characteristics of the structures in the scene. However, the dimensionality of the data is also increased. It should be noted that in our experiment all 16 classes samples are used to conduct the experiment. For the area Oats, it only has 20 samples, and 2 labeled samples are randomly chosen for training.

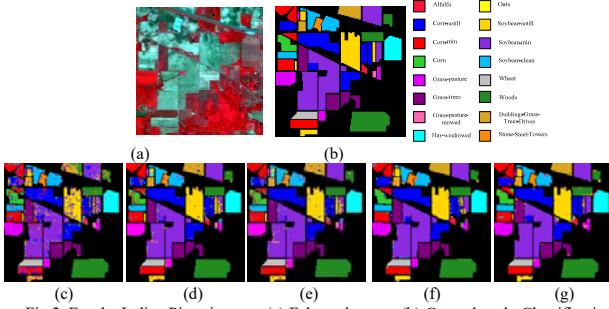


Fig.2. For the Indian Pines images: (a) False color map (b) Ground truth. Classification maps obtained by (c) SVM, (d) Spatial-Spectral-NAE, (e) SVM-CK and (f) SOMP (g) unRAE-SVM.

TABLE I  
CLASSIFICATION ACCURACY FOR INDIAN PINES  
USING DIFFERENT ALGORITHMS.

Class	Spatial-Spectral SAE	SVM	SVM-CK	SOMP	unRAE-SVM
Alfalfa	39.02	71.84	<b>85.71</b>	83.67	59.18
Corn-notill	78.37	82.52	91.48	<b>96.13</b>	95.90
Corn-min	65.60	66.86	91.88	91.48	<b>96.14</b>
Corn	72.30	60.81	80.09	<b>92.42</b>	91.94
Grass/pasture	89.89	94.17	95.08	93.30	<b>96.21</b>
Grass/trees	95.43	96.54	<b>98.51</b>	98.37	95.25
Pasture-mowed	64.00	49.58	<b>69.57</b>	50.00	45.83
Hay-windrowed	99.97	98.59	99.32	<b>99.55</b>	95.24
Oats	<b>72.22</b>	56.11	50.00	22.22	22.22
Soybeans-notill	67.89	66.36	88.75	87.39	<b>93.81</b>
Soybeans-min	79.90	84.27	96.22	97.61	<b>98.20</b>
Soybeans-clean	81.27	74.95	90.05	89.51	<b>92.41</b>
Wheat	96.20	<b>99.37</b>	97.91	98.43	98.43
Woods	96.05	96.09	99.31	<b>99.57</b>	96.39
Building-trees	55.62	61.29	96.49	<b>97.66</b>	94.44
Stone-steel	95.24	91.51	83.53	94.19	<b>96.51</b>
Overall	81.01	82.79	94.15	95.19	<b>95.52</b>
Average	78.06	78.18	<b>88.37</b>	86.97	85.51
Kappa	78.50	0.797	0.933	0.947	<b>0.949</b>

## B. ROSOS Urban Data: University of Pavia

The University of Pavia image was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS). The ROSIS sensor generates 115 spectral bands ranging from 0.43 to 0.86  $\mu\text{m}$  and has a spatial resolution of 1.3 m per pixel. The University of Pavia image includes  $610 \times 340$  pixels of nine ground-truth classes as shown in Fig. 3(a-b). By removing the 12 noisy bands, each pixel is represented by 103 bands.

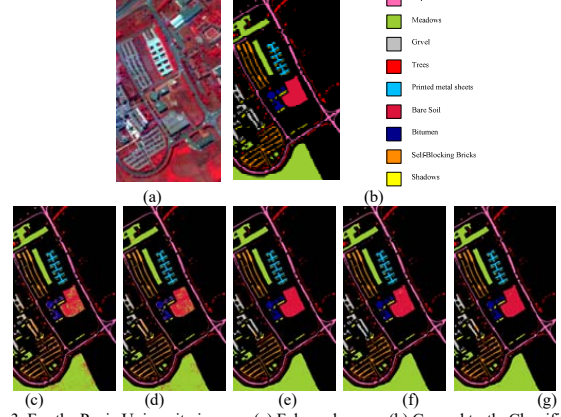


Fig.3. For the Pavia University images: (a) False color map (b) Ground truth. Classification maps obtained by (c) SVM, (d) Spatial-Spectral-NAE, (e) SVM-CK and (f) SOMP (g) unRAE-SVM.

TABLE II  
CLASSIFICATION ACCURACY FOR THE UNIVERSITY OF PAVIA  
USING DIFFERENT ALGORITHMS.

Class	Spatial-spectral SAE	SVM	SVM-CK	SOMP	unRAE-SVM
Asphalt	91.15	89.32	99.19	93.92	<b>98.81</b>
Meadows	96.82	95.82	99.82	99.88	<b>99.96</b>
Gravel	70.79	71.43	85.13	94.52	<b>96.17</b>
Trees	91.75	93.70	97.53	94.79	<b>99.04</b>
Metal Sheets	99.67	99.60	99.92	99.68	<b>99.92</b>
Bare Soil	79.61	76.49	98.73	96.96	<b>99.91</b>
Bitumen	80.08	80.31	<b>99.09</b>	97.73	97.33
Bricks	86.27	81.75	<b>97.79</b>	97.22	96.51
Shadows	96.29	91.76	<b>99.88</b>	87.79	97.11
Overall	90.93	89.48	98.53	97.36	<b>99.06</b>
Average	88.05	86.69	97.47	95.83	<b>98.31</b>
Kappa	0.879	0.857	0.981	0.967	<b>0.988</b>

In the experiment, 9% of samples per class are chosen for training and the rest are used for testing. Table II summarizes the classification results using SVM, Spatial-Spectral SAE, SVM-CK, SOMP, and unRAE-SVM. As can be seen from three indices shown in Table II, our unRAE-SVM performs the best among all the algorithms. Compared with the Indiana Pines Dataset, University of Pavia data set contains more spectral pixels and each class has relatively balanced samples. So, all the methods perform better. Our algorithm achieves around 1.7% increase compared to SOMP, and has higher increase in accuracy than the other algorithms. Especially, for the bare soil class, our unRAE achieves 99.19% in accuracy, which exceeds the other algorithms clearly.

Fig. 3(c-g) shows the visual classification results of the five algorithms. It is shown that the bare soil categories are better



classified by unRAE-SVM than the other algorithms, which is in agreement with the statistical results shown in Table II. Compared with SOMP, unRAE-SVM obtains a better classification result.

### C. Effect of Neighborhood Size

We set the neighborhood size to  $9 \times 9$ ,  $7 \times 7$ ,  $5 \times 5$ , and  $3 \times 3$  respectively to obtain the classification results on the two HSI datasets. The results are shown in Fig. 4. As we can see from Fig. 4, for both datasets, the neighborhood size with  $7 \times 7$  gets good performance. Although the overall accuracy with  $9 \times 9$  neighborhood is slightly better than  $7 \times 7$  on the University of Pavia dataset, we chose  $7 \times 7$  for both datasets to compromise the computation complexity. It is noticed that the overall accuracy will not always increase with the increase of the neighborhood size because more neighboring information unrelated to the investigated pixels are involved.

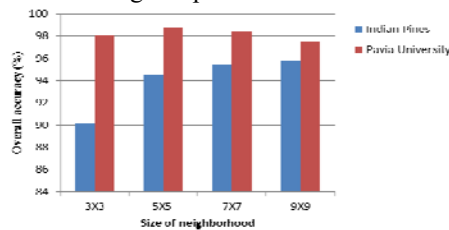


Fig. 4. Overall classification accuracy for different neighborhood sizes for Indiana Pines and University of Pavia datasets

## V. CONCLUSION

In this letter, we have proposed an unsupervised feature learning method based on the unsupervised RAE network integrated with spatial information, and have proved its effectiveness for HSI classification. Since the neighboring pixels are exploited to learn the features for the investigated pixel in the spatial neighborhood, similarity weights between the central pixel and the neighboring pixels are added to the unsupervised RAE network to learn features accurately. Moreover, our proposed unRAE algorithm learns features in an unsupervised learning manner which avoids the burden in obtaining labeled samples for HSI analysis. The model based on the unsupervised RAE for feature learning also can be extended to HSI band selection, which is our future work.

## REFERENCES

- [1] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
- [2] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral-imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [3] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, Jun. 2009.
- [4] M. Fauvel, J. Benediktsson, J. Chanussot, and J. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [5] Y. Tarabalka, J. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, 2009.
- [6] C. Rodarmel, and J. Shan, "Principal component analysis for hyperspectral image classification," *Surveying and Land Information Systems*, vol. 62, no. 2, pp. 115–123, 2002.
- [7] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote-sensing data over urban areas," in *Proc. EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–14, 2009.
- [8] X. Zhang, W. Wang, Y. Li, and L. C. Jiao, "PSO-based automatic relevance determination and feature selection system for hyperspectral image classification," *IET electronics letters*, vol. 48, no. 20, pp. 1263–1265, Oct. 2012.
- [9] X. Zhang, Y. He, N. Zhou, and Y. Zheng, "semisupervised dimensionality reduction of hyperspectral images via local scaling cut criterion," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1547–1551, Nov. 2013.
- [10] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimension gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.
- [11] F. Tsai, and J. Lai, "Feature extraction of hyperspectral image cube using three-dimension gray-level cooccurrence," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3504–3513, June. 2013.
- [12] J. J. Li, Marpu, P. R., Plaza, A., & Bioucas-Dias, J. M., "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, 2013.
- [13] Q. Wang, J. Lin, Y. Yuan, "Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 6, pp. 1279, 2016.
- [14] Y. Chen, N. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [15] X. Zhang, P. Weng, J. Feng, E. Zhang, B. Hou, "Spatial-spectral classification based on group sparse coding for hyperspectral image," *IGARSS*, pp. 1745–1748, 2013.
- [16] U. Srinivas, Y. Chen, V. Monga, and N. M. Nasraadi, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Trans. Geosci. Remote Sens.*, vol. 10, no. 3, pp. 505–509, May. 2013.
- [17] Yuan, Y., Lin, J., & Wang, Q. (2016). Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization. *IEEE transactions on cybernetics*, 46(12), 2966–2977.
- [18] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wisetraining of deep networks," in *Proc. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 2007, pp. 153–160.
- [19] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deepbelief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
- [20] K. Fukushima, "Neocognitron: A self-organizing neural network model for mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.
- [21] Y. Chen, Z. Lin, X. Zhao, G. Wang and Y. Gu, "Deep Learning-Based Classification of Hyperspectral Data", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 7, no. 6, pp. 2094–2107, 2014.
- [22] Y. Chen, H. Jiang, C. Li, X. Jia, P. Ghamisi, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks", *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016.
- [23] A. Romero, C. Gatta and G. Camps-Valls, "Unsupervised Deep Feature Extraction for Remote Sensing Image Classification", *IEEE Trans. Geosci. Remote Sens.*, Vol. 54, no. 3, pp. 1349–1352, 2016.
- [24] R. Socher and Andrew Y. Ng, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *proc. EMNLP*, 2011
- [25] R. Socher, C. Lin, Andrew Y. Ng, and C. D. Manning. "Parsing natural scenes and natural language with recursive neural networks," in *Proc. ICML*. 2011.
- [26] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Programming*, vol. 45, pp. 503–528, 1989.
- [27] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.